

# Simulated Annealing Approach to the Temperature-Emissivity Separation Problem in Thermal Remote Sensing Part One: Mathematical Background

John A. Morgan  
The Aerospace Corporation  
P. O. Box 92957  
Los Angeles, CA 90009  
john.a.morgan@aero.org

**Abstract**—The method of simulated annealing is adapted to the temperature-emissivity separation (TES) problem. A patch of surface at the bottom of the atmosphere is assumed to be a greybody emitter with spectral emissivity  $\epsilon(k)$  describable by a mixture of spectral endmembers. We prove that a simulated annealing search conducted according to a suitable schedule converges to a solution maximizing the *A-Posteriori* probability that spectral radiance detected at the top of the atmosphere originates from a patch with stipulated  $T$  and  $\epsilon(k)$ . Any such solution will be nonunique. The average of a large number of simulated annealing solutions, however, converges almost surely to a unique Maximum *A-Posteriori* solution for  $T$  and  $\epsilon(k)$ .

The limitation to a stipulated set of endmember emissivities may be relaxed by allowing the number of endmembers to grow without bound, and to be generic continuous functions of wavenumber with bounded first derivatives with respect to wavenumber.

**Keywords**—Remote Sensing; Temperature-Emissivity Separation; Surface Temperature Estimation.

## I. INTRODUCTION

THE temperature-emissivity separation (TES) problem bedevils any attempt to extract spectral information from remote sensing in the thermal infrared. A variety of methods has been proposed for handling the temperature-emissivity separation (TES) problem [1], [2]. In most of them, simultaneous LST and band emissivity retrieval depends upon specifying an emissivity value in one or more reference bands. The MODIS Land Surface Temperature (LST) algorithm [3] seeks a pair of reference channels in a part of the thermal spectrum in which the emissivity of natural surfaces displays very limited variation, and may therefore be regarded as known with good confidence. Multiband emissivities inferred on this basis are called "relative" emissivities [4]. Other algorithms of this nature include the reference channel method [5], emissivity normalization [6], temperature-independent spectral index method [7], [8] and spectral ratios [9]. The study by Li et al. [4] shows that all of these relative emissivity retrieval algorithms are closely related, and argues that they may be expected to show comparable performance. The analysis of Multispectral Thermal Imager (MTI) data [10] depends on collection of radiance from a surface with looks

at nadir and 60 degrees off-nadir, assuming a known angular dependence of emissivity, in order to balance equations and unknowns. The generalized split-window LST algorithm [11] likewise uses dual looks in a regression-law based approach. The "grey body emissivity" approach [12] exploits the slow variation of emissivity with wavelength for certain natural targets, while the physics-based MODIS LST algorithm [13] exploits observations taken at day and at night, on the assumption that band emissivities do not change over periods of a few weeks. A study with the Airborne Hyperspectral Scanner [14] compares multiple TES approaches.

We shall investigate a simulated annealing approach to the TES problem. The approach is an extension of earlier work on Bayesian TES [15], [16]. Simulated annealing cannot give a unique solution to this problem, but we shall prove that the average of a large number of simulated annealing TES solutions converges almost surely to a unique TES estimate.

This paper will concentrate on the mathematical basis of the algorithm and a proof of its convergence. A study of its performance will form the subject of a subsequent paper.

## II. BACKGROUND

Simulated annealing has traditionally been regarded as a preferred method of global solution for combinatorial optimization problems such as Traveling Salesman. In this paper, we adapt the Metropolis algorithm [17], [18], [19] to an optimization problem that lacks a unique global optimal solution: Temperature-emissivity separation. The underdetermined temperature-emissivity separation (TES) problem, notoriously [1]-[14], has a continuous infinity of solutions that yield the identical optimum value for any cost or payoff function one cares to choose.

A key part of any simulated annealing algorithm is the choice of an annealing schedule that causes the posterior probabilities to transition from nearly uniform to very tight in such a way as to evade the risk of the MAP search from converging to a local, rather than a global, optimum. The

look and feel of the justification for this approach is ergodic.

In what follows we shall mostly concern ourself with the existence of a solution to the simulated annealing TES problem, and shall simply assume that a suitable annealing schedule has been supplied. Factors that enter into the choice of annealing schedule are described in [18], [19]. Selection of the annealing schedule and sample TES retrievals will be the subject of a subsequent paper.

### III. SIMULATED ANNEALING AND THE TEMPERATURE-EMISSION SEPARATION PROBLEM

#### A. Metropolis Algorithm Search for Maximum A-Posteriori Solution

Suppose that we have in our possession prior knowledge that a target patch that forms part of the lower boundary of the atmosphere is composed of an intimate mixture of  $m+1$  spectral endmembers  $\{\epsilon_i(k)\}$  at temperature  $T$ . For later convenience, we shall require that spectral emissivities be bounded continuous functions of wavenumber with bounded first derivative with respect to wavenumber. The label  $k$  may, depending upon context, refer to wavenumber, or to a finite number of wavenumber-averaged spectral bands. Except in Section III-B.2, we shall assume the band interpretation.

The spectral mixture amounts to a mapping into a geometric  $m$ -simplex whose vertices have spectral endmembers at a stipulated temperature  $T$  for coefficients. Suppose we have  $m+1$  distinct points  $\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_m$  in  $\mathbf{R}^m$  chosen so that the vectors  $\mathbf{y}_1 - \mathbf{y}_0, \mathbf{y}_2 - \mathbf{y}_0, \dots, \mathbf{y}_m - \mathbf{y}_0$  are linearly independent. Then the set

$$K_m \equiv \sum_{i=0}^m \lambda_i \mathbf{y}_i \quad (1)$$

with

$$\lambda_i \geq 0, \forall i \quad (2)$$

and

$$\sum_{i=0}^m \lambda_i = 1 \quad (3)$$

is an  $m$ -simplex. [20] A spectral mixture with stipulated weights  $\lambda_i$  corresponds to the vector <sup>1</sup>

$$\mathbf{x} = \sum_{i=0}^m \lambda_i \mathbf{y}_i \in \mathbf{R}^m. \quad (4)$$

The *interior* of  $K_m$  is the subset of  $K_m$  for which  $\lambda_i > 0$ , that is the closure of its interior. The polyhedron of  $K_m$ , denoted  $|K_m|$ , is the set comprised of the points of  $\mathbf{x} \in K_m$

<sup>1</sup>Should the target patch contain an isothermal checkerboard mixture of end members, the weight  $\lambda_i$  is to be interpreted as the fraction of the total surface area of the patch occupied by the  $i$ -th subregion  $\lambda_i = \frac{A_i}{\sum_{j=1,n} A_j}$ . The product  $\lambda_i \epsilon_i(k)$  is thus a normalized emissivity-area product for that subregion. We defer the case of endmembers with differing temperatures to a later date. We believe it, however, to be a straightforward extension of the reasoning in this paper.

considered as a subset of  $\mathbf{R}^m$ , and is a convex compact subset of  $\mathbf{R}^m$ .

In the case  $m+1 = 3$ , a familiar example of a 2-simplex is the ternary diagram used to classify phreatic igneous rocks. The double three-component diagram used in the QAPF classification [21] scheme is a union of two 2-simplices, and is an example of a simplicial complex.

For present purposes, the  $i^{th}$  pure endmember for the  $n^{th}$  trial is assigned to the  $i^{th}$  vertex of  $K_m$

$$\mathbf{y}_i \Leftrightarrow \epsilon_i(k), 0 \leq i \leq m \quad (5)$$

with the spectral mixture corresponding to a point in the polyhedron of  $K_m$ ,

It is necessary to account for surface temperature in a somewhat different way. Let the minimum and maximum physically admissible surface temperatures be  $T_{min}$  and  $T_{max}$ , respectively. Then the temperature of our target patch is given by

$$T = (1 - \lambda_{m+1})T_{min} + \lambda_{m+1}T_{max} \quad (6)$$

with

$$0 \leq \lambda_{m+1} \leq 1. \quad (7)$$

Corresponding to  $\mathbf{x}$  introduced already, we have from Eq. (6)

$$x_{m+1} \in I^1, \quad (8)$$

the unit interval, with

$$x_{m+1} \Leftrightarrow T_n \quad (9)$$

The quantity that appears in the forward model for the  $n^{th}$  trial is

$$\langle \epsilon(k) B_k(T_n) \rangle = \sum_{i=1}^m \lambda_i \epsilon_i(k) B_k(T_n). \quad (10)$$

$B_k(T_n)$  is the (band-integrated, as needed) Planck function at temperature  $T_n$ . The parametrization of the choice  $\{T_n, \epsilon(k)\}$  in terms of the vector  $x$  is a mapping into the topological product

$$H^{m+1} \equiv I^1 \otimes |K_m| \quad (11)$$

of  $I^1$  and  $|K_m|$ . The set  $H^{m+1}$  is not a simplex, nor is it necessarily a simplicial complex. It is, however, a convex polytope, and is the convex hull of its vertices  $\mathbf{x}_i, 0 \leq i \leq m+1$ .<sup>2</sup>

We score trial mixtures by that we most wish to maximize: The posterior probability for the observed spectral radiance to originate from a surface patch with temperature  $T$  and spectral emissivity  $\epsilon(k)$ . A standard argument [15], [16] gives the posterior probability in terms of a MAXENT estimator

$$P(I | T, \epsilon, \sigma) = \exp \left[ -\frac{(I - I_{FM})^2}{2\sigma^2(T_a)} \right] \frac{dI}{\sigma(T_a)} \quad (12)$$

<sup>2</sup>Although we will not need it in what follows,  $H^{m+1}$  can be decomposed into either a simplicial complex or a union of simplices.

in terms of a forward model

$$I_{FM} = f\left(\sum_{i=1}^m \lambda_i \epsilon_i(k) B_k(T_n)\right) \Leftrightarrow f(x) \quad (13)$$

that is some function of the  $n^{th}$  trial, in each spectral bin  $k$ . We note that while the equation of transfer is linear, the dependence of its solution  $I_{FM}$  upon  $\epsilon_i(k) B_k(T_n)$  need not be. The assumed noise variance  $\sigma^2$  is shown as having a formal dependence upon a parameter, the "annealing temperature"  $T_a$ , which governs the annealing schedule for the search for a Maximum A-Posteriori solution. The joint posterior probability in  $J$  spectral bands is proportional to

$$P(\{I_k\} | T, \epsilon, \sigma) = \prod_{k=1}^J \exp\left[-\frac{(I_k - I_{FM}(k))^2}{2\sigma^2(T_a)}\right] \frac{dI}{\sigma(T_a)} \quad (14)$$

If radiance  $I_k$  in each of  $J$  bands originating from a patch on the Earth's surface has been detected at the top of the atmosphere (TOA), the posterior probability that the surface patch is at a temperature  $T$  given prior knowledge  $K$  is given by Bayes' theorem as

$$P(T, \epsilon_i(k) | \{I_k\}, K) = P(T, \epsilon(k) | K) \frac{P(\{I_k\} | T, \epsilon_i(k), K)}{P(\{I_k\} | K)}. \quad (15)$$

The noise variance is assumed known and the functional dependence of probabilities upon  $\sigma_i$  is omitted. The prior probability  $P(\{I_i\} | K)$  for the radiances  $\{I_k\}$  has no dependence upon  $T$  and for our purposes may be absorbed into an overall normalization. [22] Equation (15) is evaluated with aid of the prior probability for the surface to be at temperature  $T$  and have spectral emissivity  $\epsilon(k)$ , given available knowledge  $K$  [15],

$$P(T, \epsilon(k) | K) dT \propto \prod_k d\epsilon(k) \frac{dT}{T}. \quad (16)$$

$P(T, \epsilon_i(k) | \{I_i\}, K)$  is the conditional probability for the hypothesis that the surface temperature is  $T$ , and the spectral emissivity  $\epsilon_k$ , given observed radiances  $\{I_i\}$  and prior knowledge  $K$ .

Each trial is thus scored according to the joint posterior probability for observed spectral radiance  $I_i$  to result from surface temperature  $T$  and spectral emissivity  $\epsilon_k$ ,

$$p_n = P(T_n, \epsilon(k) | \{I_i\}, K) \equiv p_n(x) \quad (17)$$

where  $x$  stands for  $\{\mathbf{x}, x_{m+1}\}$ . Thus, in going from the  $(n-1)^{th}$  to the  $n^{th}$  trial, the  $n^{th}$  candidate mixture is selected by Monte Carlo draw and  $p_n$  for the new trial is compared to  $p_{n-1}$  for the last one. The probability that it is accepted is [17], [19]

$$P = \begin{cases} 1 & \text{if } p_n/p_{n-1} \geq 1 \\ P(T_a) & \text{otherwise} \end{cases} \quad (18)$$

where the probability  $P(T_a)$  of taking a downward step in  $p_n$  is determined by the annealing schedule. The dependence of  $P(T_a)$  on the annealing schedule is symbolized by

the annealing "temperature"  $T_a$  which is taken to decrease systematically during the MAP search. The actual form  $P(T_a)$  takes in practical calculations is determined empirically.

## B. Convergence

### B.1 Spectral mixtures comprising a finite number of end-members

We now examine the question of convergence. Corresponding to the sequence of  $m$ -simplices  $K_m$  as the number of trials  $n$  increases without bound is a sequence of trials  $\{T_n, \epsilon_i(k)\}$  with associated loci  $\{x\} \in H^{m+1}$ .

As a closed bounded subset of  $\mathbf{R}^{m+1}$ ,  $H^{m+1}$  is a compactum. Therefore, as  $n \rightarrow \infty$ , the sequence of trials  $x$  contains a convergent subsequence, whatever the value of  $m$ . Correspondingly, the sequence of posterior probabilities likewise has a convergent subsequence that, by construction, tends to the maximum value of the posterior probability, i.e., to a MAP solution for  $T$  and  $\epsilon(k)$ .

Consider the map  $x' = \Phi(x)$  given by

$$\Phi(x) = \begin{cases} x' & \text{if } p_n(x') - p_n(x) \geq 0 \\ x & \text{otherwise.} \end{cases} \quad (19)$$

The mapping Eq. (19) gives the action of the Metropolis algorithm according to Eq. (18) at sufficiently late times in the annealing schedule that a transition to a state of decreased posterior probability occurs rarely; in the limit, almost never. We have noted that at a sufficiently late point in the annealing schedule, trials that decrease the posterior probability Eq. (17) will become infrequent. We may elide any such trials without affecting the convergence of the subsequence, which then takes the form

$$x_{n+1} = \Phi(x_n) \quad (20)$$

For all  $n$  greater than some  $M$ , convergence of the subsequence implies the Cauchy condition

$$d(x_n, x_{n+1}) = d(x_n, \Phi(x_n)) < \epsilon, \quad (21)$$

(with the Euclidean norm supplying a suitable metric for finite  $m$ ) so that

$$x \rightarrow \Phi(x). \quad (22)$$

The mapping Eq. (19) generates a sequence of trials  $x$  for which  $p_n$  is nondecreasing. By Zorn's Lemma, the set comprised of all admissible trials  $x$  has at least one element with a maximal value of  $p_n$ . We note that maximizing  $p_n$  also maximizes the information-theoretic entropy by Eq. (12). According to the usual statement of the Second Law, the state of maximum entropy is one of thermodynamic equilibrium. We may therefore, in a nod to Refs. [17] and ([25] both, call the limit Eq. (22) an equilibrium point.

We note that, in the limit, Eq. (22) amounts to a fixed point of Eq. (19). Ordinary fixed-point theorems are inapplicable to Eq. (19) because it is neither continuous nor semicontinuous: It can map an open set  $\in H^{m+1}$  to a singleton  $x'$ . We can, however, adapt a celebrated construction introduced by Nash [25] to prove the existence of a fixed point of an equivalent self-mapping.

In fact, we shall prove a somewhat stronger result. Consider

$$\phi_\alpha = \max(0, p_n(x_\alpha) - p_n(x)). \quad (23)$$

for stipulated  $x$ . The function  $\phi$  is continuous in the mixture  $x_\alpha$ . Define the mapping  $N : x \rightarrow x'$  by

$$x' = \frac{x + \sum_\alpha \phi_\alpha x_\alpha}{1 + \sum_\alpha \phi_\alpha}, \quad (24)$$

where the index  $\alpha$  is taken to run over members of any finite set of admissible trials  $x_\alpha$  in the execution of the Metropolis algorithm. (One may think of the collection of all sequences  $x_\alpha$  in ensemble-theoretic terms.) Suppose that  $x'$  is a fixed point under Eq. (24). In Eq. (24) some values of  $\alpha$  correspond to choices for  $\{T_i, \epsilon_i(k)\}$  for which the posterior probability does not increase:

$$p_n(x_\alpha) - p_n(x) \leq 0. \quad (25)$$

For these values of  $\alpha$ ,

$$\phi_\alpha = 0. \quad (26)$$

If the choice  $x$  is fixed under the mapping  $N$  in Eq. (24), then the contribution to  $x'$  from any  $x_\beta$  must not decrease; therefore,  $\phi_\beta = 0, \forall \beta$ , lest the denominator in  $\Phi$  exceed unity. Put another way, no other choice of  $\{T_i, \epsilon_i(k)\}$  can increase the posterior probability. But that is the definition of an equilibrium point.

If, on the other hand, an equilibrium point  $x$  maximizes the posterior probability Eq. (17), every  $\phi_\alpha$  vanishes, so that  $x$  is a fixed point.

Equation (24) is continuous and maps points  $x$  into a convex compactum  $\subset \mathbf{R}^{m+1}$ . A fixed point

$$x = N(x) \quad (27)$$

exists according to the Brouwer fixed-point theorem that, by construction, maximizes the *a-posteriori* probability of  $x$ .

The mapping Eq. (19) generates a sequence of trials  $x$  for which  $p_n$  is nondecreasing and gives the maximal value of  $p_n$  in the limit, while Eq. (27) demonstrates the existence of a trial  $x^*$  for which  $p_n$  cannot be made greater. In view of the ensemble-theoretic freedom to choose  $x_\alpha$ , we may identify the limit in Eq. (22) with the fixed point in Eq. (27). Therefore, a convergent subsequence of annealing trials exists that tends to an equilibrium point. Moreover, Eq. (27) demonstrates that the annealing search can, in principle, find  $x^*$  in a finite number of trials. We conclude that, granted a suitable annealing schedule, there exists at least one convergent sequence of trials that tends to MAP surface temperature and spectral emissivity estimates consistent with observed spectral radiances  $I_k$ .

## B.2 Arbitrary spectral emissivities

The search algorithm just described assumes that the emissivity  $\epsilon(k)$  is describable by a mixture of a finite set

of spectral end members. While the spectral mixture characterization of  $\epsilon(k)$  is of interest in its own right, it may be considered a stronger hypothesis than is strictly desirable. In particular, it seems intuitively reasonable that the simulated annealing approach to TES should work just as well-and admit a simpler algorithmic realization-by using trials with randomly chosen spectral emissivities, rather than by seeking a spectral mixture from a predetermined set of endmembers.

In fact, it is possible to reduce the case of search using arbitrary  $\epsilon(k)$  for trials to an extension of the analysis in the preceding section by allowing the number of endmembers  $m$  to grow without limit for each trial  $n$ . Instead of self-mappings into a single polytope with fixed  $m$ , we consider a sequence of  $H^{m_n+1}$  as  $m_n \rightarrow \infty$ , for each  $n$  in the annealing schedule.

The connection between arbitrary  $\epsilon(k)$  and a spectral mixture with whose endmembers are allowed to grow without limit is easily seen. If we chose  $x_i, 0 \leq i \leq m_j$  from a set of randomly chosen endmembers  $\epsilon_i(k)$ , it is clear that any random  $\epsilon(k)$  can be constructed as a spectral mixture of other random endmembers. By induction: A single endmember  $\epsilon_1(k)$  trivially reproduces an arbitrary  $\epsilon(k)$  if it is chosen so  $\epsilon_1(k) = \epsilon(k)$ . Suppose that any  $\epsilon(k)$  equals a spectral mixture of  $j$  suitably chosen random  $\epsilon_i(k)$ . Then, by the inductive hypothesis for  $j = 2$ , it is possible to replicate any other  $\epsilon(k)$  by a mixture of some spectral emissivity of  $j$  endmembers and a  $(j+1)^{st}$  random endmember.

We proceed by constructing the polytope for an arbitrary number of spectral endmembers. The polyhedron of the m-simplex  $K_m$  with unit diameter may be circumscribed by an m-sphere of radius [23]

$$r \leq \sqrt{\frac{m}{2(m+1)}}. \quad (28)$$

$|K_m|$  is thus a subset of the topological product of  $n$  replicas of the unit interval  $[0, 1]$

$$|K_m| \subset I^m \equiv [0, 1] \otimes [0, 1] \cdots [0, 1] \quad (29)$$

Every m-simplex is thus contained within the topological product of a countable infinity of replicas of the unit interval  $[0, 1]$  [24]

$$|K_m| \subset I^\infty \equiv [0, 1] \otimes [0, 1] \cdots, \quad (30)$$

as is every convex polytope

$$H^{m+1} \subset I^\infty. \quad (31)$$

All the polyhedra  $|K_m|$  and polytopes  $H^{m+1}$  are compact, and by Tychonoff's theorem, the set  $I^\infty$  which circumscribes every  $|K_m|$  and  $H^{m+1}$  is likewise sequentially compact.

$I^\infty$  is homeomorphic to the *Hilbert cube*. The Hilbert cube

$$\mathcal{H} \equiv [0, 1] \otimes [0, \frac{1}{2}] \cdots [0, \frac{1}{n}] \cdots \quad (32)$$

is a subset of a Hilbert space with the  $l_2$  norm.<sup>3</sup>  $I^\infty$  is therefore a *complete* space: The sequence of trials, by sequential compactness of the Hilbert cube, and sets homeomorphic to it, possesses a convergent subsequence whose limit is, by completeness of  $I^\infty$ , an element of that space. The limit of the convergent subsequence of m-polyhedra is likewise contained within  $I^\infty$ .

The treatment of convergence in Section III-B.1 requires modification when the number of spectral endmembers is allowed to grow without limit. At each  $n$ , the self-mapping  $N$  and function  $\phi(x)$  are applied to  $H^{m_n+1}$  as before to give existence of an equilibrium fixed point. The simplex dimension  $m_n$  is allowed to grow without bound, however. The resulting sequence of equilibrium points  $\in H^{m_n+1}$  possesses a convergent subsequence  $\subset H^{m_n+1}$  for each value of  $n$ .

With that caveat, as  $n \rightarrow \infty$  the sequence  $\{x_n\}$  converges to a MAP estimate of  $T$  and an arbitrary  $\epsilon(k)$  by the same reasoning used in the previous section. In consequence, the limiting equilibrium point of the convergent sequence  $\Phi(x_n) = x_n$  as  $m_n \rightarrow \infty$  will tend to an estimate of the MAP value for  $\{T, \epsilon_i(k)\}$  for any admissible  $\epsilon(k)$ . We conclude that a suitable simulated annealing search will converge to an arbitrary spectral emissivity that gives a MAP estimate of  $\{T, \epsilon_i(k)\}$ .

### C. Uniqueness

Whatever the dimensionality of the spectral endmember parameterization of emissivity, sequential compactness guarantees existence of a convergent subsequence of trials. In practice, we must expect that there will be more than one such sequence. The nonuniqueness of solutions to the TES problem suggests that there will be a continuous infinity of possible trials  $\{T_n, \epsilon_i(k)\}$  that yield any stipulated value for the posterior probability. In any realizable search strategy, however, we need only contend with a countable set of convergent subsequences. Amongst these there will be one for which the posterior probability is greatest.<sup>4</sup> This will be the closest approach to the Maximum A-Posteriori solution achieved by simulated annealing. In the nature of things, more than one convergent subsequence may be expected to exist that yields this same maximal estimate, with the same asymptotic annealing temperature  $T_a^\infty$ . We ignore all subsequences except these maximal ones.

In References [15] and [16] expectation values for  $T$  and  $\{\epsilon(k)\}$  over the the posterior probability Eqn. (15) were shown to give good estimates for physical surface temperatures and emissivities. We claim that the mean of a large number of subsequences that converge to the limiting MAP value will tend to the expectation values for  $T$  and  $\{\epsilon(k)\}$  with respect to Eqn. (15).

The MAXENT estimator is constructed from the posterior probability of noise power in a spectral bin. For the sake of simplicity we assume identical noise power in each

bin.<sup>5</sup> A fully annealed MAP estimate may be thought of as an individual Bernoulli trial drawn from the likelihood function for  $\{T_n, \epsilon_i(k)\}$ . By construction, all such trials are independent and identically distributed with bounded expectation values.<sup>6</sup>

Let

$$\bar{T} = \frac{1}{N} \sum_{i=1}^N T_i \quad (33)$$

and

$$\overline{\epsilon(k)} = \frac{1}{N} \sum_{i=1}^N \epsilon_i(k) \quad (34)$$

be the means of MAP surface temperature and spectral emissivity taken over over  $N$  convergent subsequences. Suppose the covariance matrix  $\Sigma$  of the trials to be nonsingular. We invoke the multivariate Central Limit Theorem to conclude the mean values converge weakly to the multivariate Gaussian distribution:

$$\sqrt{N} \begin{pmatrix} \bar{T} - \langle T \rangle \\ \overline{\epsilon(1)} - \langle \epsilon(1) \rangle \\ \vdots \\ \overline{\epsilon(m)} - \langle \epsilon(m) \rangle \end{pmatrix} \rightsquigarrow \mathbf{N}_m(0, \Sigma). \quad (35)$$

Reliance on the mixing hypothesis in the form given by Eq. (10), however, brings with it the concern that the relevant covariance matrix might be singular. In that event, the strong law of large numbers [30], [31] ensures

$$\bar{T} \xrightarrow{a.s.} \langle T \rangle \quad (36)$$

and

$$\overline{\epsilon(k)} \xrightarrow{a.s.} \langle \epsilon(k) \rangle, \quad (37)$$

but without giving estimated variances of the mean values, such as come with Eq. (35).

To the extent that the estimator used in the simulated annealing search is zero-mean error, we conclude the estimates yield accurate values for the physical values of  $T$  and  $\{\epsilon(k)\}$ . As the spectral weights  $\mathbf{x}$ , lying as they do between zero and unity, possess bounded moments, this conclusion applies to the limiting mean values of  $\{T_n, \{x_m^n\}\}$  as well.

### D. Arbitrarily fine spectral resolution

It is worth considering briefly the limiting case of infinite spectral resolution for  $\epsilon(k)$ . In this section only,  $k$  refers to wavenumber. We shall admit as endmembers any continuous function  $\epsilon_i(k)$  on a compact interval  $[k_1, k_2] \in \mathbf{R}$ , with

$$0 \leq \epsilon_i(k) \leq 1, \quad (38)$$

and bounded first derivative

$$\left\| \frac{d\epsilon_i(k)}{dk} \right\|_{sup} < W \quad (39)$$

<sup>5</sup>This assumption is inessential and may be relaxed.

<sup>3</sup>For any finite dimensional subspace of  $\mathcal{H}$  however, we may still take the Euclidean norm when choosing a metric.

<sup>4</sup>One may appeal to Zorn's lemma again at this point, if desired.

<sup>6</sup>Moments over Eqn. (15) are bounded despite bad behavior of the Jeffreys prior at  $T=0$ , because of the rapid decay of the exponentials away from the MAP solution, as *L'Hôpital's* rule demonstrates.

on that interval.

Consider

$$f_{m_n}(k) = \epsilon(k) - \sum_{j=0}^{m_n} \lambda_j \epsilon_j(k), \quad (40)$$

where, as before, as  $m_n$  increases without bound, an admixture of arbitrary  $\epsilon_j(k)$  is included in the spectral mixture. A standard argument shows that Eqs. (38) and (39) imply the spectral emissivities and Eq. (40) are all members of an equicontinuous set. By the Arzelà-Ascoli lemma, as  $m_n \rightarrow \infty$  there is a subsequence of trial emissivity spectra

$$\epsilon^t(k) = \sum_{j=0}^{m_n} \lambda_j \epsilon_j(k) \quad (41)$$

for which an  $M$  exists such that for  $m_n > M$  and for any positive  $\delta$

$$|f_{m_n}(k)| < \delta \quad (42)$$

uniformly on  $[k_1, k_2]$ . Thus, any physically admissible spectral emissivity  $\epsilon(k)$  may be approximated arbitrarily well by a suitable spectral mixture of an unlimited number of end members.

We may regard the spectral mixture  $x$  as an upper function on  $[0, 1]$  and take for a norm the Lebesgue measure in the limit  $m_n \rightarrow \infty$ , with associated metric

$$d(x, y) = \int d\mu(x - y). \quad (43)$$

A mixture  $x$  then becomes a vector in a Banach space. As  $n \rightarrow \infty$  the sequence of limiting fixed points under the self-mapping  $\Phi$  has a convergent subsequence  $x_n$  which, again, satisfies the Cauchy condition

$$d(x_n, x_{n+1}) < \epsilon \quad (44)$$

The argument from the Cauchy property of the convergent subsequence of  $x_n$  to the conclusion that the subsequence converges to a MAP equilibrium point likewise follows much as before. As  $n \rightarrow \infty$

$$0 < p(x_{n+1}) - p(x_n) < \epsilon, \quad (45)$$

so that the convergent subsequence of annealing trials, again, tends to an equilibrium point.

#### IV. DISCUSSION

In Section III-B we proved convergence of simulated annealing searches for candidate MAP TES solutions. In Section III-C we argued that the average of a large number of these candidate MAP solutions converges almost surely to a unique estimate of surface temperature and spectral emissivity that, given a forward model leading to an unbiased estimator for  $T$  and  $\{\epsilon_k\}$ , closely approximates the true values of these quantities.

The motivation for seeking the Nash equilibrium analogy came from the realization that spectral mixing theory [27], [28], [29] amounts to the use of mappings into a simplex,

and that iterative choice of weights in the TES problem according to an annealing schedule amounts to a self-mapping into a convex polytope. The analogy with spectral mixing theory, however, is incomplete: Spectral unmixing, in either its reflective or thermal variants, generally appends and extra end member called "virtual dark" or "virtual cold", used to accommodate the effect of contamination from noise and sensor artifacts. One may see the value of an extra end member by recalling that the highest order components in a principle components decomposition of a multivariate dataset tends to be dominated by noise and artifacts that do not correlate with the physical content of lower-order components. The role of a virtual garbage end member will be discussed in Part Two.

A special case of great interest is the situation in which one seeks evidence that spectral radiance sensed at the top of the atmosphere (TOA) contains evidence for the presence of a specific spectral component. This problem may be addressed by use of spectral mixtures comprised of the desired spectral component together with generic continuous end members whose number is allowed to grow without bound. The analysis of Section III-B.2, however, cannot simply be modified by appending an endmember corresponding to the desired emissivity spectrum, as any completely random choice of  $\epsilon(k)$  may well be correlated with the stipulated spectrum and so introduce a spurious admixture of that endmember into the analysis. On the other hand, in a physical mixture of spectral endmembers such correlations might in fact occur naturally. Without going into details, we offer some thoughts on this problem.

One way to proceed is to randomly select  $\epsilon_i(k)$  for the background in such a way that the otherwise arbitrary background emissivity spectra all lie in the null space of the stipulated endmember. The choice of background  $\epsilon(k)$  is thus made in much the same way as an empirical orthogonal basis set is selected.

Depending on the intended application, another way to handle this problem might be to find some way of marginalizing on the set of arbitrary background  $\epsilon(k)$  in the calculation of the posterior probability Eq. (17). If performed by the same sort of stochastic sampling typical of simulated annealing, this calculation would resemble a numerical approximation to the Feynman-Kac formula.[32]; in essence, a Monte Carlo path integral calculation.

#### V. CONCLUDING REMARKS

The practical utility of the mathematical development in this paper may be questioned. We address briefly two possible concerns.

While convergence of the algorithm has been proved to our satisfaction, we have no equally satisfactory estimates of the rate of convergence, with the consequence that the choice of annealing schedule remains a matter of trial-and-error. In response to this concern, the availability of massively parallel computation made possible by the ready availability of cheap GPU arrays means that massive processing requirements need not preclude the use of

a resource-hungry algorithm if that algorithm can provide performance not attainable by other approaches. Part Two of this study will address these issues.

Another legitimate concern is that the spectral emissivity of natural ground covers in the wild will seldom be known to the level of accuracy found in Ref. [26]. While true in general, this concern has not dissuaded other researchers from relying upon spectral unmixing.

The simulated annealing approach to TES by spectral unmixing does, however, offer something that other TES algorithms do not: By construction, it gives (in the limit) the unique best estimate in a Maximum *A-Posteriori* sense, for the remote determination of surface temperature and spectral emissivity of a patch of ground that is known to be comprised of a spectral mixture of a stipulated set of spectral end members.

## REFERENCES

- [1] Dash, P., F.-M. Göttsche, F.-S. Olesen, and H. Fischer, "Land surface temperature and emissivity estimation from passive sensor data: theory and practice-current trends," *Int. J. Remote Sensing*, vol. 23, pp. 2563-2594, 2002
- [2] Li, Z.-L., H. Wu, N. Wang, S. qiu, J. A. Sobrino, Z. Wan, B.-H. Tang, and G. Yan, "Land surface emissivity retrieval from satellite data," *International Journal of Remote Sensing* **34**, pp. 3084-3127, 2013
- [3] Wan, Z.-M., *MODIS Land-Surface Temperature Algorithm Theoretical Basis Document*, Institute for Computational Earth System Science, University of California, Santa Barbara, 1999
- [4] Li Z.-L., F. Becker, M. P. Stoll, and Z. Wan, "Evaluation of Six Methods for Extracting Relative Emissivity Spectra from Thermal Infrared Images," *Rem. Sens. Env.*, vol. 69, pp. 197-214, 1999
- [5] Kahle, A. B., and R. E. Alley, "Separation of Temperature and Emittance in Remotely Sensed Radiance Measurements," *Rem. Sens. Env.*, vol. 42, pp. 107-111, 1992
- [6] Kealy, P. S., and S. J. Hook, "Separating Temperature and Emissivity in Thermal Infrared Multispectral Scanner Data: Implications for Recovering Land Surface Temperatures," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 1155-1164, 1993
- [7] Petitcolin, F., and E. F. Vermote, "Land Surface Reflectance, Emissivity and Temperature from MODIS Middle and Thermal Infrared data," *Rem. Sens. Env.*, vol. 83(1-2), 112-134, 2002
- [8] Li Z.-L., and F. Becker, "Feasibility of Land Surface Temperature and Emissivity Determination from AVHRR Data," *Rem. Sens. Env.*, vol. 43, pp. 67-85, 1993
- [9] Watson, K., "Spectral Ratio Method for Measuring Emissivity," *Rem. Sens. Env.*, vol. 42, pp. 113-116, 1992
- [10] Borel, C. C., and J. . Szymanski, "Physics-based Water and Land Temperature Retrieval," in *Handbook of Science Algorithms for the Multispectral Thermal Imager*, B. W. Smith, Ed., Los Alamos National Laboratory and Savannah River Technology Center, 1998
- [11] Wan, Z.-M., and J. Dozier, "A generalized split-window algorithm for retrieving land-surface temperature from space," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 892-905, 1996
- [12] Barducci, A., and I. Pippi, "Temperature and emissivity retrieval from remotely sensed images using the 'Grey body emissivity' method," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 681-695, 1996
- [13] Wan, Z.-M., and Z.-L. Li, "A physics-based algorithm for land-surface emissivity and temperature from EOS/MODIS data," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 980-996, 1997
- [14] Sobrino, J. A., J. C. Jiménez-Muñoz, P. J. Zareco-Tejada, G. Sepulcre-Cantó, Eduardo de Miguel, "Land surface temperature derived from airborne hyperspectral scanner thermal infrared data," *Rem. Sens. Env.*, vol. 102, pp. 99-115, 2006
- [15] Morgan, J. A., "Bayesian Estimation for Land Surface Temperature Retrieval: The Nuisance of Emissivities," *IEEE Transactions on Geosciences and Remote Sensing*, **43**, pp. 1279-1288, 2005
- [16] Morgan, J. A., "Comparison of Bayesian land surface temperature algorithm performance with Terra MODIS observations," *International Journal of Remote Sensing* **32**, 2012, p. 8139; see also arXiv:0909.4461v3, 2011
- [17] Metropolis N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.* **21**, pp. 1087-1091, 1953
- [18] Kirkpatrick, S., C. D. Gelatt, M. P. Vecchi, "Optimization by Simulated Annealing," *Science* **220**, pp. 671-680, 1983
- [19] Press, W., Flannery, B., Teukolsky, S., and Vetterling, W., *Numerical Recipes in C*, Cambridge University Press, Cantab., 1988, pp. 326-334
- [20] Rotman, J. J., *An Introduction to Algebraic Topology*, Springer-Verlag, New York, 1988, pp. 31-38
- [21] Le Bas, M. J., and A. L. Streckeisen, "The IUGS systematics of igneous rocks", *Journal of the Geological Society, London*, **148**, pp. 825-833, 1991
- [22] Bretthorst, L., "Bayesian Spectrum Analysis and Parameter Estimation," in Berger, J., S. Fienberg, J. Gani, K. Krickenberg, and B. Singer, Eds, *Lecture Notes in Statistics*, Springer-Verlag, New York, 1988, vol. 48
- [23] Blumenthal, L. M., and G. E. Wahlin, "On the Spherical Surface of Smallest Radius Enclosing a Bounded Subset of n-Dimensional Euclidean Space," *Bulletin of the American Mathematical Society* **47**, pp. 771-777, 1941
- [24] Milnor, J., "The Geometric Realization of a Semi-Simplicial Complex," *Annals of Mathematics* **65**, pp. 357-362, 1957
- [25] Nash, J. F., "Non-Cooperative Games," *Annals of Mathematics* **54**, pp. 286-295, 1951
- [26] Salisbury, J. W., L. S. Walter, N. Vergo, and D. M. D'Aria, *Infrared (2.1-25  $\mu$ ) Spectra of Minerals*, Johns Hopkins University Press, Baltimore, 1992
- [27] Johnson, P. E., M. O. Smith, and J. B. Adams, "Quantitative Analysis of Planetary Reflectance Spectra with Principal Components Analysis," *J. Geophys. Res.* **90**, pp. C805-C810, 1985
- [28] Adams, J. B., M. O. Smith, and P. E. Johnson, "Spectral Mixture Modeling: A New Analysis of Rock and Soil Types at the Viking Lander 1 Site," *J. Geophys. Res.* **91**, pp. B8098-B8112, 1986
- [29] Adams, J. B., M. O. Smith, and P. E. Johnson, "Correction to Spectral Mixture Modeling: A New Analysis of Rock and Soil Types at the Viking Lander 1 Site," *J. Geophys. Res.* **91**, p. B10,513, 1986
- [30] van der Waart, A. W., *Asymptotic Statistics*, Cambridge University Press, Cantab., 1998
- [31] Artstein, Z., and R. A. Vitale, "A Strong Law of Large Numbers for Random Compact Sets," *The Annals of Probability* **3**, pp. 879-882, 1975
- [32] Kac, M., *Probability and Related Topics in the Physical Sciences*, Interscience Publishers, London, Chapter IV, 1959

John Morgan is a Senior Engineering Specialist in the Sensing and Exploitation Department at The Aerospace Corporation, where his duties include studies of spacecraft sensor system performance and remote sensing phenomenology. He has a BS from Caltech, and MS and Ph.D. degrees from Rice University, in physics.